Proposal for Using NLP Interchange Format for Question Answering in Organizations

Majid Latifi

Department of Software, Universitat Politècnica de Catalunya – BarcelonaTech(UPC), Barcelona, Spain
mlatifi@lsi.upc.edu

Abstract. The growth of technology and sciences has greatly influenced the area of management and decision-making procedures, and has dramatically changed the decision-making processes in different levels, both quantitatively and qualitatively. Knowledge management plays a vital role in supporting enterprise learning, since it facilitates the effective collective intellect of the enterprise. Different methods for user-friendly knowledge access have been developed previously. The most sophisticated ones provide a simple text box for a query which takes Natural Language (NL) queries as input. Question Answering (QA) system is playing an important role in current search engine optimization. Natural language processing technique is mostly implemented in QA system for asking user's question and several steps are also followed for conversion of questions to query form for getting an exact answer. Query languages have complex syntax, requiring a good understanding of the representation schema, including knowledge of details like namespaces, class and property names. In this research we proposed an model to implement Conceptual Question Answering and Automatic Information Inferences for the enterprise's operational knowledge management in ontology-based learning organization.

Keywords: Enterprise Ontology, Learning Organization, Question Answering(QA), Information Inference, NLP.

1 Introduction

Retrieval and extraction processes - for enterprise management and decision-making - have gained an excessive importance as the mass of data and information stored in various resources increases. Knowledge is considered a key factor for enterprise prosperity at present and future. Knowledge management is an integrated, systematic process that applies a suitable combination of information technologies and human cooperation in order to identify, manage and share the information capitals. In addition, it both includes the explicit and implicit knowledge of the staff and it applies various and extensive methods to retrieve, store and share knowledge in a certain enterprise.

The application of "Semantic Web" technologies to learning processes is receiving an increasing attention from the perspective of facilitating the selection, delivery and tailoring of learning experiences. But most of the current approaches are centered on

the final interaction of the learner with the "learning objects" provided for him/her, neglecting the organizational perspective. From the viewpoint of an organization, the application of Semantic Web technologies should be motivated by the improvement of learning-oriented mechanisms, including both cultural and structural aspects, and considering the ideal of achieving a state of continuous improvement in learning behavior. Such an approach to achieving a "semantic learning organization" gives a complementary perspective to existing "educational Semantic Web" propositions [2]. A main need for the semantic enterprise model is one which extracts and displays the enterprise semantics.

Most knowledge bases provide facilities for querying through the use of some formal language such as SPARQL or SeRQL. However, these have a fairly complex syntax, requiring a good understanding of the data schema and being prone to errors due to the need to type long and complicated URIs. These languages are homologous to the use of SQL for interrogating traditional relational databases and should not be seen as an end user tool[13].

The obvious solution to these problems is to create some additional abstraction level that provides a user friendly way of generating formal queries. It may be possible to infer from this information for the machine so that we can carry out the decision-making and planning procedures in enterprise processes through automatic inference.

2 Statement of the Problem and Related Work

A basic method to transform an organization into a learning organization is to apply knowledge management within the organization. By facilitating the process of creating and sharing knowledge, and through providing positive working environments and effective rewarding systems, knowledge management accelerates enterprise learning and helps the enterprise adjust itself to today's rapid changes and hence survive in pace with these changes[9]. By using ontology, we can identify the meanings related to a domain, an enterprise or a society or even determine these meanings within different societies in details as desired [3]. In Ontology-based QA system, the knowledge based data, where the answers are sought, has a structured organization. The question-answer retrieval of ontology knowledge base provides a convenient way to obtain knowledge for use, but the natural language need to be mapped to the query statement of ontology. Accessing structured data such as that encoded in ontologies and knowledge bases can be done using either syntactically complex formal query languages or complicated form interfaces that require expensive customization to each particular application domain.

Probably due to the extraordinary popularity of search engines such as Google, people have come to prefer search interfaces which offer a single text input field where they describe their information need and the system does the required work to find relevant results. While employing this kind of interface is straightforward for full text search systems, using it for conceptual search requires an extra step that converts

the user's query into semantic restrictions like those expressed in formal search languages. Following are discussed some examples of such query interfaces.

CLOnE[9], presents a controlled language for ontology editing and a software implementation, based partly on standard NLP tools, for processing that language and manipulating an ontology. The input sentences are analyzed deterministically and compositionally, which the software consults in order to interpret the input's semantics; this allows the user to learn fewer syntactic structures since some of them can be used to refer to either classes or instances, for example. A repeated-measures, task-based evaluation has been carried out in comparison with a well-known ontology editor.

The Controlled Language for Ontology Editing (CLOnE) allows users to design, create, and manage information spaces without knowledge of complicated standards (such as XML[1], RDF[2] and OWL[3]) or ontology engineering tools. It was implemented as a simplified natural language processor that allows the specification of logical data for semantic knowledge technology purposes in normal language. CLOnE is designed either to accept input as valid or to reject it and warn the user of his errors; because the parsing process is deterministic, the usual IE performance measures (precision and recall) are not relevant.

QACID [10] is based on collection of queries from a given domain which are analyzed and grouped as clusters and those are manually annotated using SPARQL queries. Each query is considered as bag of words, mapping between words in NL queries into KB by using string distance metrics. SPARQL generator replaces the ontology with instances mapped for original NL query. It is domain specific and the performance depends on the types of questions collected in domain.

ONLI (Ontology Natural Language Interaction) [11] is a natural language question answering system used as front-end to the RACER reasoner and to nRQL, RACER's query language. ONLI assumes that the user is familiar with the ontology domain and works by transforming the user's natural language queries into nRQL. No details are provided regarding the effort required for re-purposing the system.

QAAL [12] surveys different types of question answering system based on ontology and semantic web model with different query format. For comparison, the types of input, query processing method, input and output format of each system and the performance metrics with its limitations was analyzed and discussed. There are basically three types of question classification methods available. Those are machine learning approaches, knowledge based approach and template based approach. In QAAL system is used template based approach for fast retrieval of answer. If the question is already asked in that system, the retrieval takes place within question template table, otherwise matching is performed using Graph Matching Algorithm and uses Spread Activation Algorithm for query matching with the ontology.

---

[1] eXtensible Markup Language

[2] Resource Description Framework

[3] Web Ontology Language

QuestIO [13] system has a natural language interface for accessing structured information, that is domain independent and easy to use without training. It brings the simplicity of Google's search interface to conceptual retrieval by automatically converting short conceptual queries into formal ones, which can then be executed against any semantic repository. The QuestIO application is open-domain (or customizable to new domains with very little cost), with the vocabulary not being predefined but rather automatically derived from the data existing in the knowledge base. The system works by converting NL queries into formal queries in SeRQL. It was developed especially to be robust with regard to language ambiguities, incomplete or syntactically ill-formed queries, by harnessing the structure of ontologies, fuzzy string matching, and ontology-motivated similarity metrics. It works by leveraging the lexical information already present in the existing ontologies in the form of labels, comment and property values.

PANTO [14] model a Portable nAtural laNguage inTerface to Ontologies which accepts input as natural language form and the output is in SPARQL query. It is based on triple model in which parse tree is constructed for the data model using the off-the-shelf Standford parser. Logic rules are applied for natural language queries as negation, comparative and superlative form. For mapping WordNet and String metric algorithms are used. The parse tree forms the intermediate representation as Query Triples Form. Then PANTO converts Query Triples form into OntoTriples form which are represented as entities in ontology.

OntoTriples are finally interpreted as SPARQL form. The performance of PANTO is analyzed by using FMeasure type. At the maximum 88.05% Precision is achieved for Geography domain with tested queries. So this system helps bridge the gap between the real world users with the semantic web based on logic model.

AquaLog [15] is capable of learning the user's jargon in order to improve his experience by the time. Their learning mechanism is good in a way that it uses ontology reasoning to learn more generic patterns, which could then be reused for the questions with similar context. In this system two major models are used as Linguistic Component which is used to convert the NL questions into Query-triple format and Relation Similarity Service (RSS) which takes Query Triple form into Onto-Triple form. The data model is triple like {Subject, Predicate, Object} type. The Performance is based on Precision, Recall and also failure types are referred separately. At average 63.5 % of successive answers are retrieved from ontology with closed domain environment.

QASYO [16] is a sentence level question-answering system that integrates natural language processing, ontologies and information retrieval technologies in a unified framework. It accepts queries expressed in natural language and YAGO [18] ontology as inputs and provides answers drawn from the available semantic markup which combining several powerful techniques in a novel way to make sense of NL queries and to map them to semantic markup. Semantic analysis of questions is performed in order to extract keywords used in the retrieval queries and to detect the expected answer type. In the QASYO model there are 4 phases: question classifier, linguistic component, query generator and query processor which characterizing it´s architec-

ture as a waterfall model. One NL query gets translated into a set of intermediate, triple-based representations, query-triples, and then these are translated into ontology-compatible triples.
The whole QA process is composed of two consecutive phases: question analysis and answer retrieval. This model requires both an evaluation of its query answering ability. Another extension is to provide information about the nature and complexity of the possible changes required for the ontology and the linguistic component.

Knowledge management system includes methods for obtaining or gathering information, organizing, distributing and sharing information among the staff in an organization. In this research, the potential role of the Semantic Web Technology as a driver for advanced learning organizations and Question Answering system is focused on providing access to the information stored in a KB by means of natural language queries.

3 Research Objectives

The current research is aimed to show that using standard NLP tools, ontology and informal to formal semantic query model proposed in the current research can establish a relationship among various sectors including duties, activities, resources and information structure of a certain enterprise so that managerial requirements can be desirably met through semantic modeling. As a result, we may have a better chance of using this information for the managers and the users through conceptual queries on the information system of the enterprise. In attention to the actual state of semantic web technology and NLP, the recommended path for organizations that are committed to the view of a learning organization is that of first addressing infrastructural elements. Such infrastructures can be considered as the study and provision of the ontologies for each aspect of the semantic learning organization. Therefore, how can we improve knowledge management in enterprises through an appropriate selection based on ontology?. Also, how can we respond to the managerial requirements of the enterprises from simple decisions to strategic ones and how can we perform automatic extraction of the information?. Consequently, the following objectives are followed in parallel with works carried out previously:

1. Conceptual framework for the notion of a semantic learning organization with using semantic search model instead of using normal keyword search model is provided.
2. Designing and presenting a method to translate user´s semantic queries into well-defined queries using the results of NLP Interchange Format (NIF) to answer the semantic questions.
3. The necessity to be robust and ability to deal with all kinds of input including ungrammatical text, sentence fragments, short queries, etc.

4 Scope of Activity

4.1 Learning Organization Ontology

The existing organizational architecture is faced with a semantic shortage between humans and systems for having a precise and general understanding of them, which in turn causes communication problems between humans and systems or vice versa. These problems prohibit the materialization of the organizations in an assembled and concordant form with other organizations [7]. Our goal is not only to design a 'conceptual' ontology model but also to implement it as an operational ontology. This approach, mainly favored by the research community, may be beneficial for integrating the domain ontology model with an inference engine for the language. Trying to match the users´ requests by providing appropriate formal commands is faced with restrictions, and thus making such semantic query by programmers is demanding, time consuming and inefficient.

4.2 Translating Natural Language Questions into Well-defined Queries

There is technically too complicated to represent and comprehend the domain for a domain expert who has little knowledge in the well-defined queries. More importantly, from a practical point of view, there is no publicly known robust engine to manage a large KB with practical performance. On the other hand, we should increase the machines' capability in understanding the organizational structure (Intelligent-making). To this end, having analyzed the existing concepts in the scope of knowledge management of the learning organizations, we reckon the significance of the information capitals of an enterprise through an ontology-based method. Answering to semantic questions will help increase the capability of learning organizations.

The growing interest in Semantic Web applications and need to translate natural language question into a machine-readable format create many uses for such applications. It is implemented as a natural language processor that allows the specification of logical data for semantic knowledge technology purposes in normal language, but with high accuracy and reliability. The components are based on NLP Interchange Format(NIF) with using statistical machine translation method.

5 Modelling of Conceptual Question Answering Method in Learning Organizations

We designed an initial model to implement Conceptual Question Answering and Automatic Information Inferences for the enterprise's operational knowledge management in learning organization. To achieve this goal, we evaluate the SPARQL and SeRQL languages for semantic search. In [5] is shown an application of SPARQL-DL query language to natural language processing, more especially as a rule engine to use within a semantic parser. As shown, the use of such formalism for this task has several advantages including the straightforward conversion of a typed dependency graph

in an ontology. In Fig. 1, the general model of our proposed system is represented. It has the following modules.

- **Query Parsing and Analysis:** In this phase, the analytical operation of the question is found out. This Analysis is responsible for Natural Language Processing (NLP). It is a technique to identify the type of a question, type of an answer, subject, verb, noun, phrases and adjectives from the question. Tokens are separated from the question and the meaning is analyzed and the reformulation of question is sent to the next stage. The input is concerted into Natural Language and is implemented using word segmentation algorithm. In word segmentation algorithm the input query from the user is divided as keywords which is further subdivided and searched in knowledge base to get correct answers.

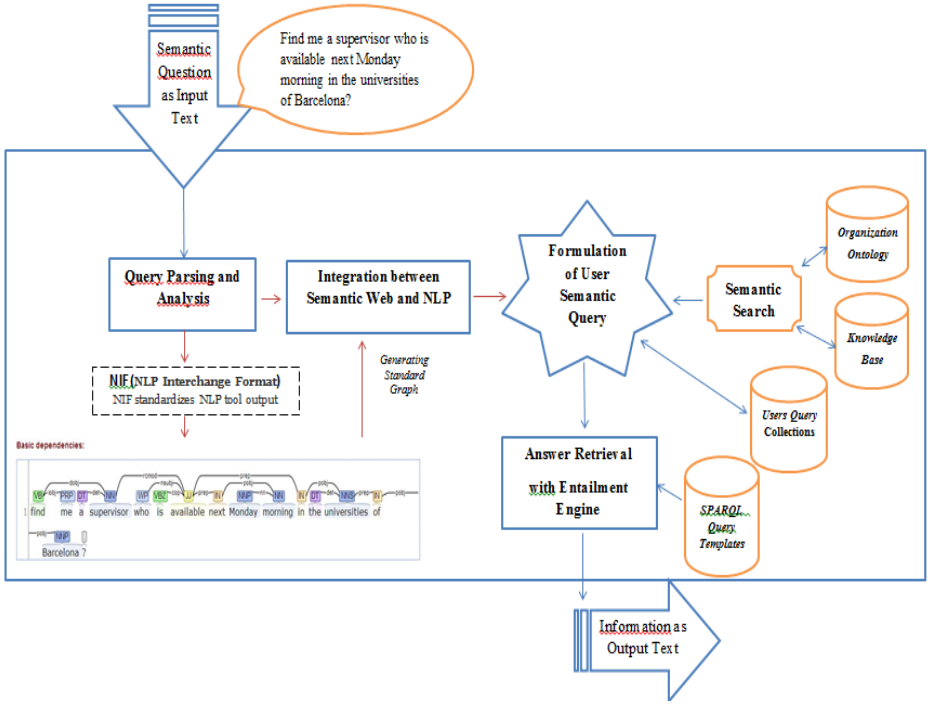

Fig. 1. Suggested Model for Semantic Question Answering

- **Integration between Semantic Web and NLP:** The tools available nowadays for Natural Language Processing can achieve very good results on many complex tasks such as the parsing of a sentence. An NLP Interchange Format for integrating NLP applications is presented by [6]. NIF addresses weaknesses of centralized integration approaches by defining an ontology-based and linked-data aware text annotation scheme. The NLP Interchange Format (NIF) is an RDF/OWL-based format that aims to achieve interoperability between NLP tools, language resources and annotations. The core of NIF consists of a vocabulary, which allows to represent strings as RDF resources. By being directly based on RDF, Linked Data and

ontologies, NIF also comprises crucial features such as annotation type inheritance and alternative annotations, which are cumbersome to implement or not available in other NLP frameworks [17].

- **Regenerating of Semantic Query:** According to the user's choice, the formulation of query is generated with the help of YAGO[18] and WordNet [19] which are implemented as semantic matching model.
- **Semantic Search:** At next stage, the Search is carried out using Conceptual Graph Matching algorithm which is the best technique. All the sentences in repository are framed as conceptual graph and the given question is also framed as conceptual graph. The matching of question CG with given CG are checked out using CG matching algorithms and the result us displayed at front-end of the our system. Graph patterns are important concepts in semantic search. RDF model is organized and graph patterns are used to formulate and encode constraint queries for locating sub graph in RDF network.
- **Graph Matching in Ontology:** Conceptual Graph acts as an intermediate language for mapping natural language questions and assertions to a relational database. Conceptual Graph (CG) contains concept, concept relation and argument. It is a graph which represents logic based on semantic model of artificial intelligence and existential graphs. Resource Description Framework (RDF) is a framework which contains triple syntax to express annotations as subject, predicate and object. Information resources are commonly represented as uniform Resource Identifiers (URIs). URIs are described by RDF. RDF triples are visualized as directed labeled graph in which subject; objects are represented as nodes and predicates as arcs.
- **Searching Ontology Nodes:** Semantic Search Algorithm is based on Conceptual Graph form of user query and domain ontology. In [8] Spread Activation is a method for searching the nodes in ontology as in semantic manner. It exploits relations between nodes in ontology. Nodes may be terms, class, object etc. Relations are labeled directed or weighted manner. SA algorithm creates initial nodes that are related to the content of the user's query and assign weights to them. After that, nodes will activate with different nodes on ontology by some rules.
- **Template based Approach:** There are basically three types of question classification methods are available. Those are machine learning approaches, knowledge based approach and template based approach. In this research we use template based approach for fast retrieval of answer. If the question is already asked in that system, the retrieval get from question template table form, otherwise matching is performed using matching algorithm.
- **Answer Retrieval with Entailment Engine:** This part of the system is based on an entailment engine. This module uses entailment techniques to infer semantic deductions between a users´query collections and the SPARQL query collections included in the formulation of user semantic query previously obtained. This process allows the system to associate new incoming queries with their corresponding SPARQL expressions in order to retrieve the answer sought from the RDF database.

6 Conclusions

The main undertaking of the current contribution is to present ongoing work in facilitating learning organizations and their use of ontology-based tools by striving to translate natural language queries into well-defined queries and retrieving exact answers, which in turn can be executed in the framework presented here . A model was introduced to automatically convert semantic query to formal query in a bid to provide answers for conceptual question and to infer information from organizational knowledge base.

Answers are retrieved from ontology using semantic search approach interoperability for NIF components, web services and question-to-query algorithm is evaluated in our system for analyzing performance evaluation. Finally performance of question answering system of getting exact result can be improved by using semantic search methodology to retrieve optimum answers from organizational ontology model.

Acknowledgments

I would like to thank our software department (LSI) from KEMLG research group in Polytechnic University of Catalonia (UPC). Especially, I would like to thank Dr. Miquel Sànchez-Marrè for his helpful comments and guidance. I acknowledge the financial support of the Generalitat de Catalunya through the AGAUR agency for Consolidated Research Groups. This support (2009SGR 1365) was granted to the Knowledge Engineering & Machine Learning group (KEMLG).

References

1. Latifi, M., Khotanlou, H., Latifi, H.: An Efficient Approach Based On Ontology to Optimize the Organizational Knowledge Base Management for Advanced Queries Service. In: 3rd IEEE International Conference on Communication Software Networks(ICCSN), ISBN: 978-1-61284-485-5, pp. 269 – 273 (2011)
2. Sicilia, M., Lytras, M.: The Semantic Learning Organization. In: The Learning Organization, Vol. 12 Iss: 5, pp.402 – 410 (2005)
3. Daconta, M., C., Smith, K., T., Obrst, L., J.: The Semantic Web: A Guide to the Future of XML, Web Services, and Knowledge Management. John Wiley & Sons, USA (2003)
4. Aggestam, L.: Learning Organization Or Knowledge Management: Which Came First, The Chicken Or The Egg?. In: Information Technology and Control, vol 35, No.3 (2006)
5. Vitucci, N., Arrigoni Neri, M., Tedesco, R., Gini, G.: Semanticizing syntactic patterns in NLP processing using SPARQL-DL queries. Politecnico di Milano, Dipartimento di Elettronica e Informazione Via Ponzio 34/5, 20133 Milano, Italy (2012)
6. Hellmann, S., Lehmann, J., Auer, S.: NIF: An ontology-based and linked-data-aware NLP Interchange Format. http:// svn.aksw.org (2012)
7. Kang, D., Lee, J., Choi, S., Kim, K.: An ontology-based Enterprise Architecture, Expert Systems with Applications, pp.1456-1464 (2010)
8. Suchal, J., Caching spreading activation search. Slovak University of Technology(2007)

9. Funk, Adam, et al.: CLOnE: Controlled language for ontology editing. The Semantic Web, Springer Berlin Heidelberg, pp.142-155 (2007)
10. Fernandez, O., R. Izquierdo, S. Ferrandez and J.L. Vicedo, Addressing ontology-based question answering with collections of user queries. Inform. Proces. Manage., 45: 175-188. DOI: 10.1016/j.ipm. (2008)
11. Shamima Mithun, Leila Kosseim, V.H.: Resolving quantifier and number restriction to question owl ontologies. In: Proceedings of The First International Workshop on Question Answering (QA2007), Xian, China (2007)
12. Kalaivani, S., and K. Duraiswamy, Comparison of Question Answering Systems Based on Ontology and Semantic Web in Different Environment. Journal of Computer Science 8.9, pp: 1407-1413 (2012)
13. Tablan, V., Damljanovic, D, Bontcheva, K, A Natural Language Query Interface to Structured Information, Springer-Verlog Berlin Heidelburg, ESWC 2008, pp. 361-375 (2008)
14. Wang, C., M. Xiong, Q. Zhou and Y. Yu, PANTO: A portable natural language interface to ontologies. Proceedings of the 4th European Semantic Web Conference, (ESWC' 07), Publication post of DBLP, pp: 473-487(2007)
15. Lopez, V., Motta, E.: Ontology driven question answering in AquaLog. In: NLDB 2004 (9th International Conference on Applications of Natural Language to Information Systems), Manchester, UK (2004)
16. Moussa, Abdullah M., and Rehab F. Abdel-Kader,: QASYO: A Question Answering System for YAGO Ontology. International Journal of Database Theory and Application 4.2 (2011)
17. Schierle, M.: Language Engineering for Information Extraction. Phd thesis, University at Leipzig, Leipzig (2011)
18. F. M. Suchanek, G. Kasneci and G.Weikum,: YAGO: A Core of Semantic Knowledge Unifying WordNet and Wikipedia. In Proceedings of 16th International World Wide Web Conference (IW3C2), pp. 697-706 (2007)
19. Miller, George A.: WordNet: a lexical database for English. Communications of the ACM 38.11, pp. 39-41 (1995)